LinkedIn

Traffic Shift - Avoiding Disasters at Scale

Michael Kehoe
Staff SRE
LinkedIn

Anil Mallapur
SRE
LinkedIn

Overview

- LinkedIn Architectural Overview
- Fabric Disaster Recovery
- Questions

World’s largest professional network

467+ million members

200+ Countries

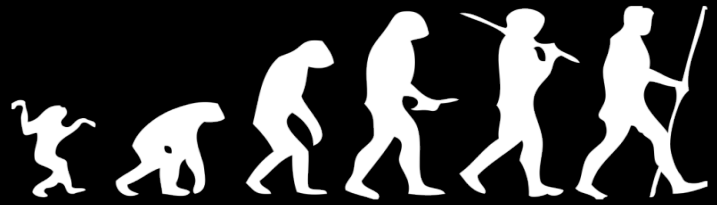
Who are we ?

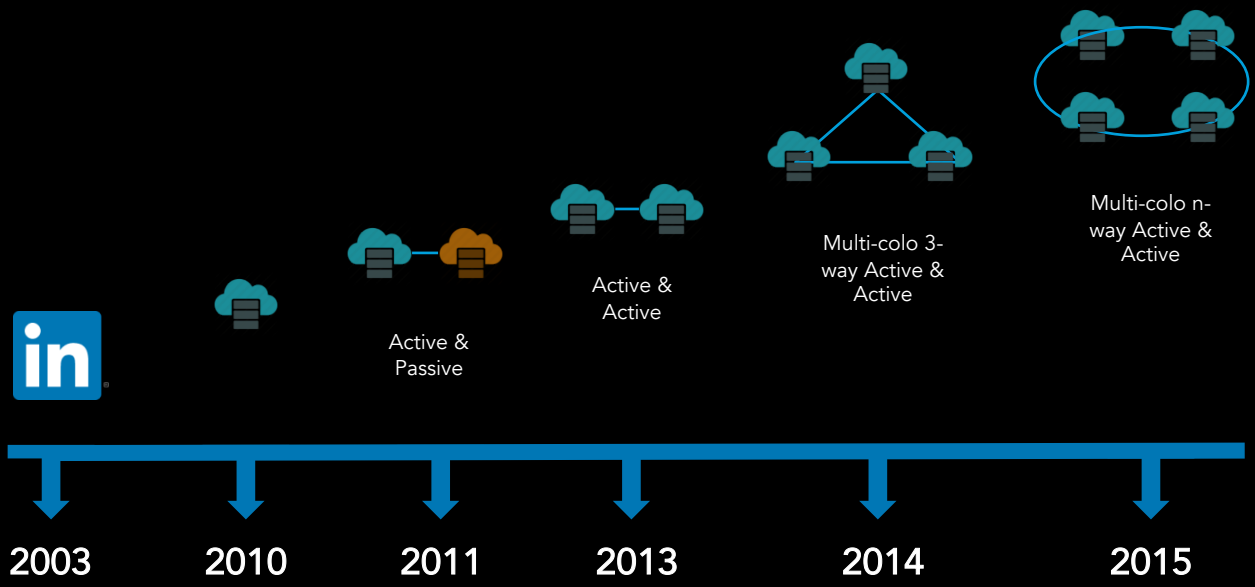
Production-SRE team at LinkedIn

- Assist in restoring stability to services during site critical issues
- Developing applications to improve MTTD and MTTR
- Provide direction and guidelines for site monitoring
- Build tools for efficient site issue troubleshooting, issue detection & correlation

Terminologies

Fabric/Colo	Data Center with full application stack deployed
PoP/Edge termination)	Entry point to LinkedIn network (TCP/ SSL
Load Test	Planned stress testing of data centers



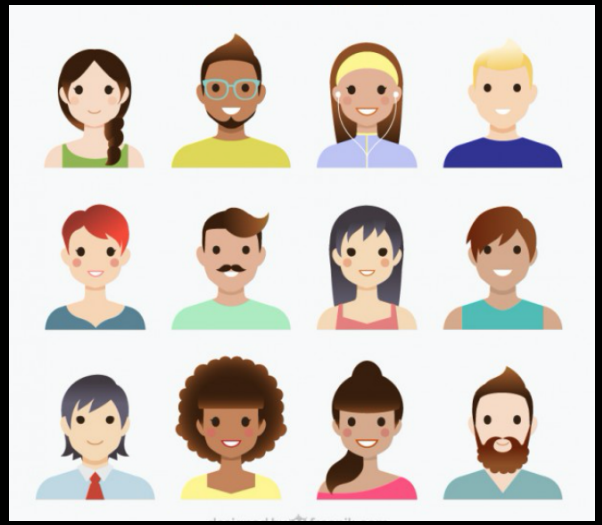

Active & Passive
Active & Active
Multi-colo 3-way Active & Active
Multi-colo n-way Active & Active
2003
2010
2011
2013
2014
2015

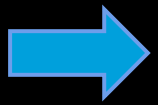
2017

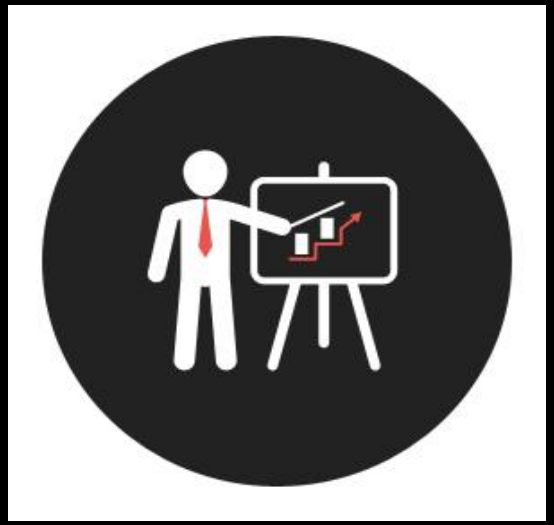
4 Data Centers

13 PoPs

1000+
services







What are Disasters ?

Service Degradation

Infrastructure Issues

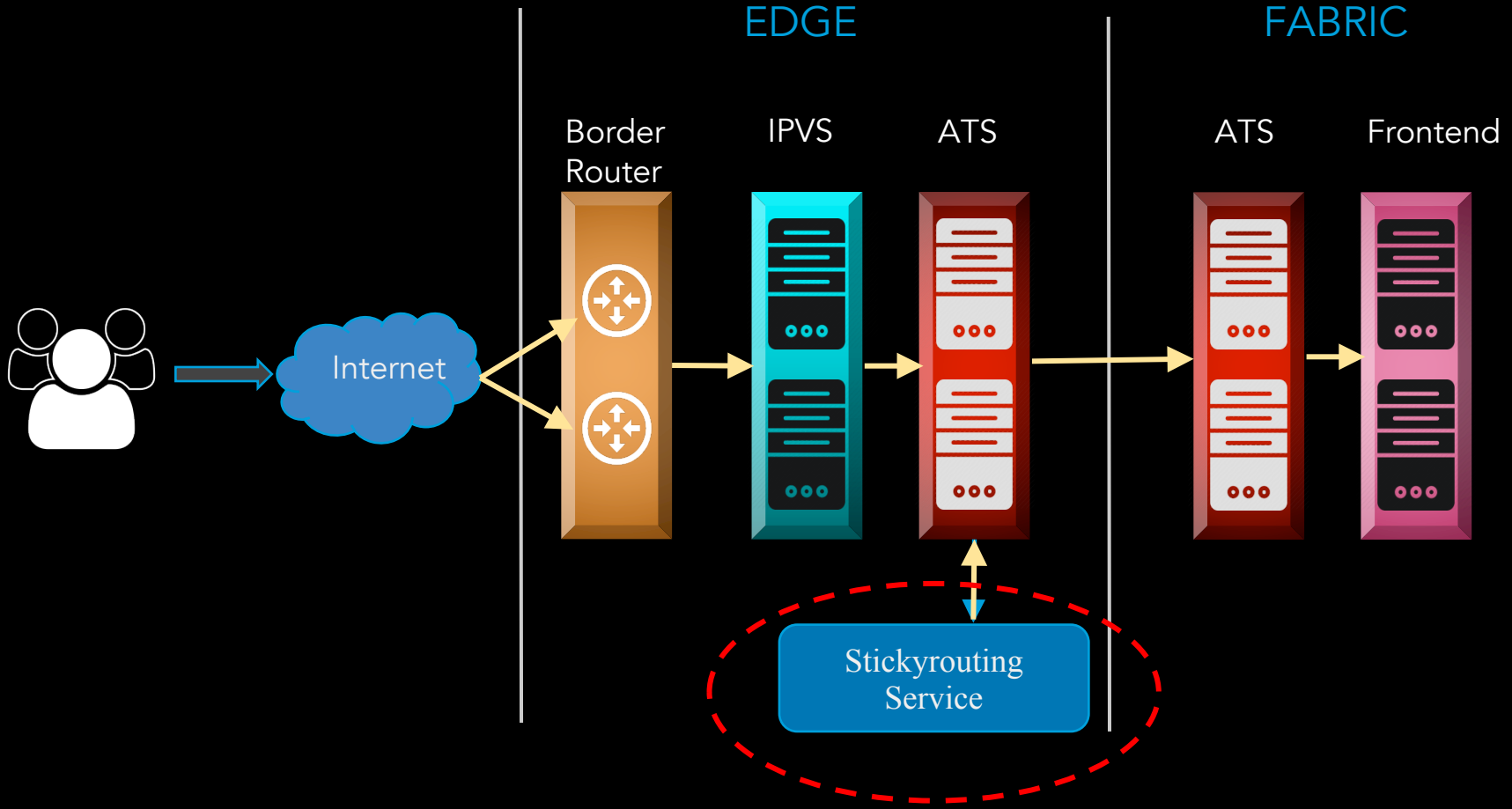
Human Error

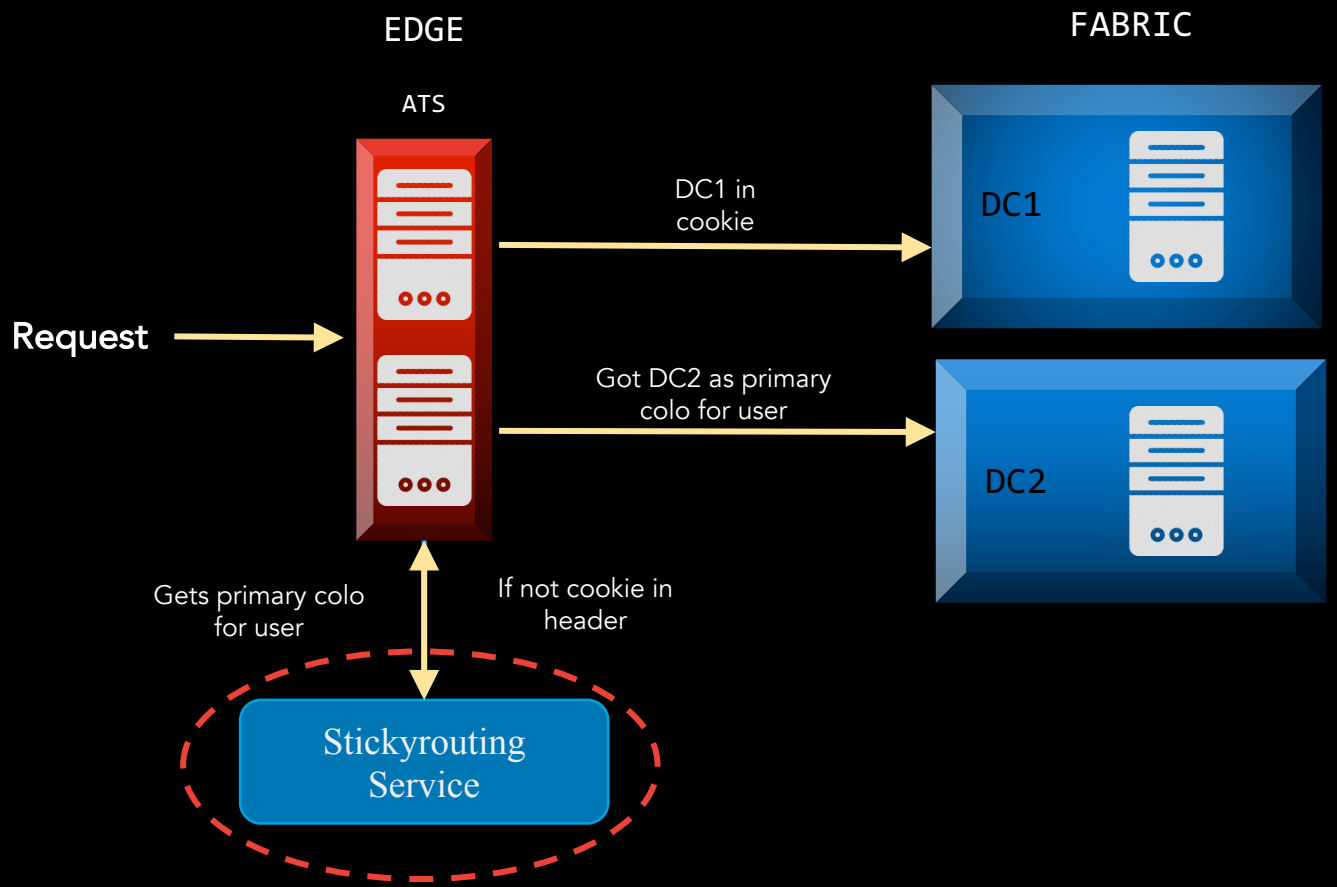
Data Center on Fire

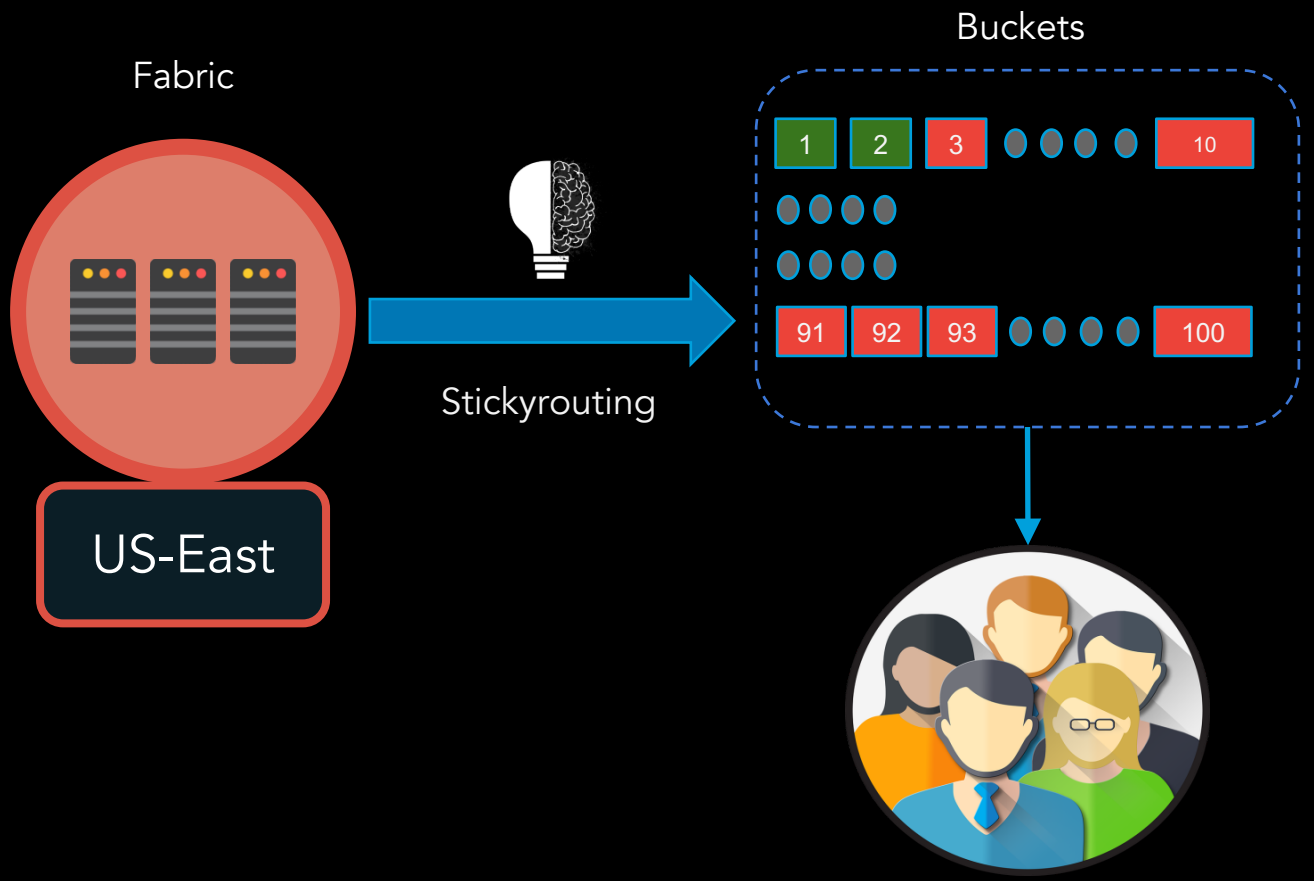
One solution for all disasters

Traffic Shift - Reroute user traffic to different datacenters without any user interruption.

Whaaaat ?


EDGE
FABRIC
Border Router
IPVS
ATS
ATS
Frontend
Internet
Stickyrouting Service

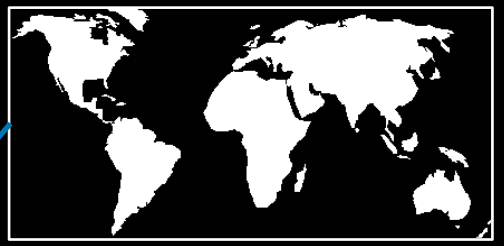

EDGE
FABRIC
ATS
DC1 in cookie
DC1
Request
Got DC2 as primary colo for user
DC2
Gets primary colo for user
If not cookie in header
Stickyrouting Service


Buckets
Fabric
1
2
3
10
91
92
93
100
Stickyrouting
US-East

How StickyRouting assigns users to a colo ?

Capacity of Fabric

Offline job
to assign
colo to users



Geographic
distance to
users

Advantages of sticky routing

Less latency for users

Store data where it’s necessary

Provides precise control over capacity allotment

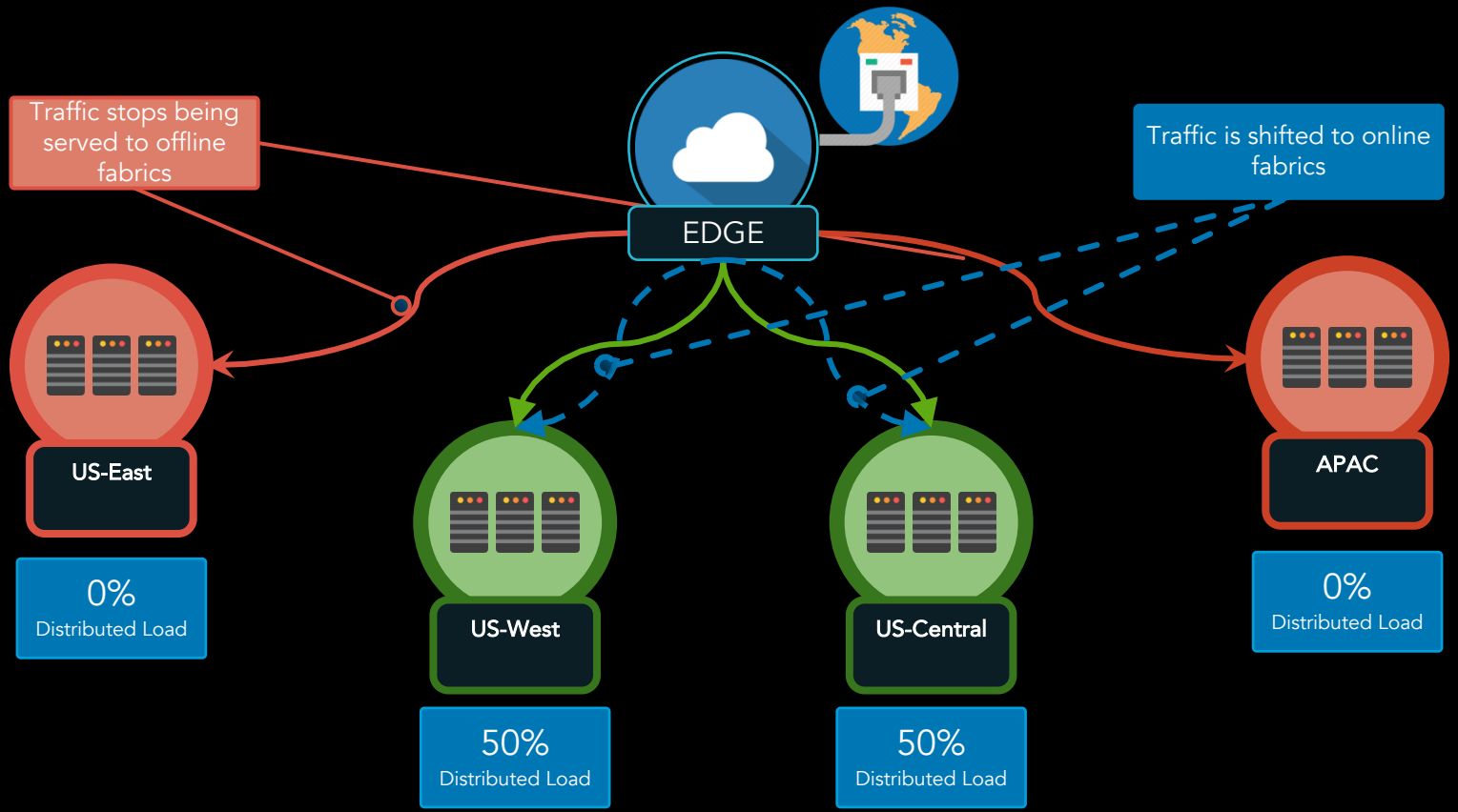
When to TrafficShift ?

Impact Mitigation

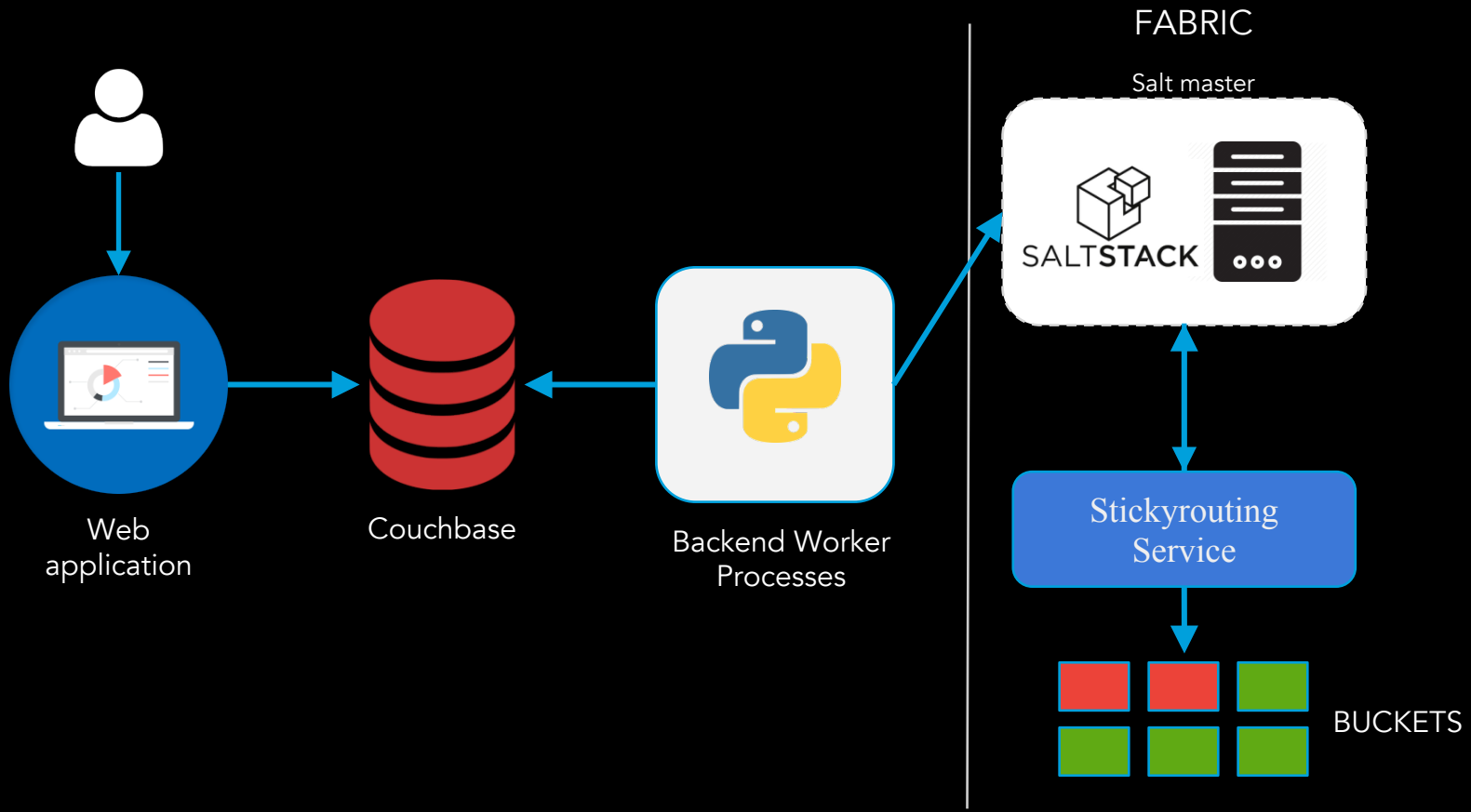
Planned Maintenance

Stress Test

Site Traffic and Disaster Recovery



TrafficShift Architecture



What is Load Testing ?

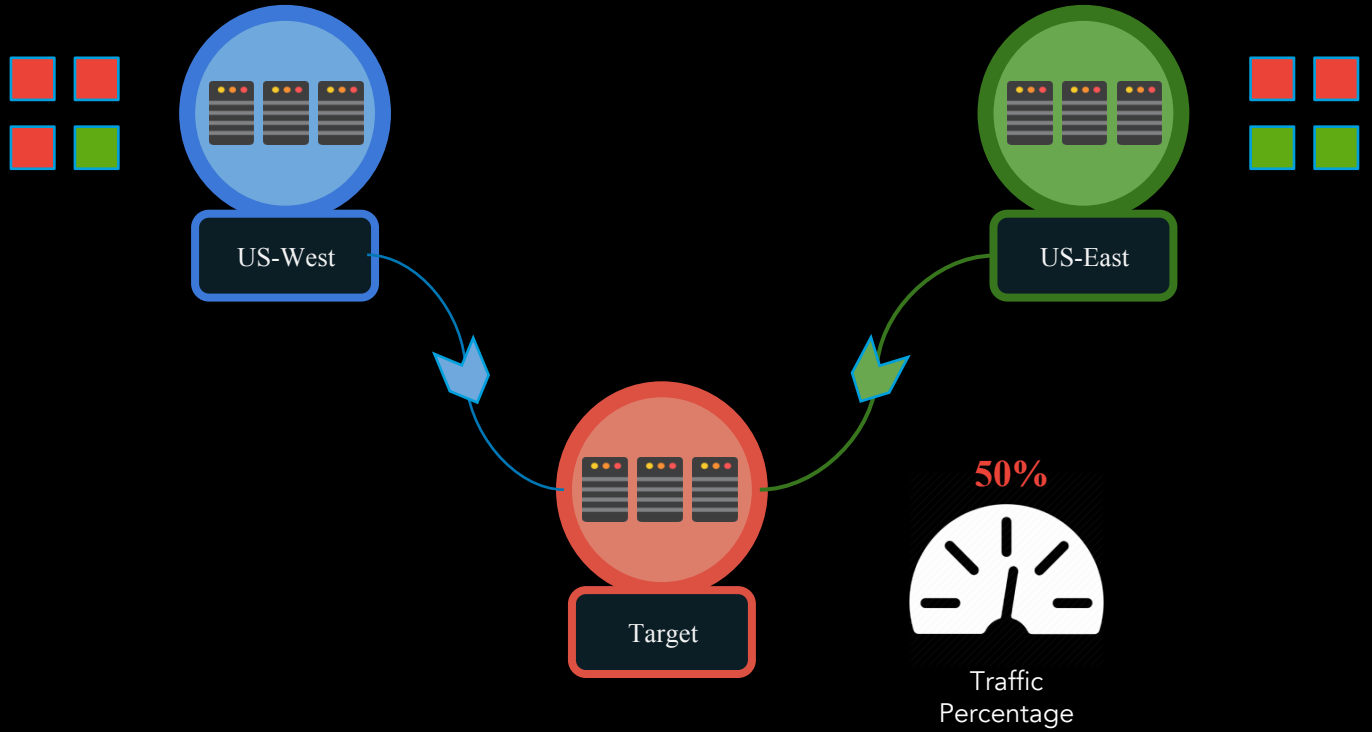
USW → Target Data Center

3 times a week

Peak hour traffic

Fixed SLA

Load Testing



Benefits of Load Test

Capacity Planning

Leverage production traffic to stress test services

Identify bugs in production

Confidence in Disaster Recovery

Big Red Button

Kill switch (No Kidding)

Failout of a datacenter and PoP in less than 10 minutes

Minimal user impact

Key Takeaways

- Design infrastructure to facilitate disaster recovery
- Stress test regularly to avoid surprises
- Automate everything to reduce time to mitigate impact

Questions?

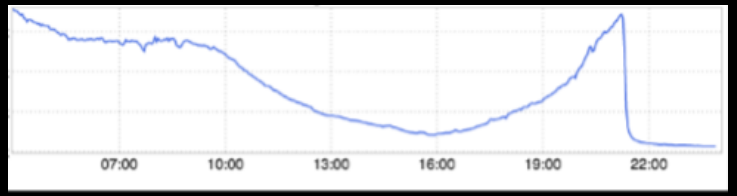
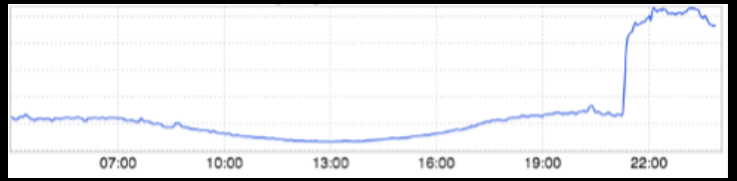
LinkedIn

Edge Failout

Edge Presence

LinkedIn's PoP Architecture

- Using IPVS - Each PoP announces a unicast address and a regional anycast address
  - APAC, EU and NAMER anycast regions
- Use GeoDNS to steer users to the 'best' PoP
- DNS will either provide users with an anycast or unicast address for www.linkedin.com
  - US and EU members is nearly all anycast
  - APAC is all unicast

LinkedIn's PoP DR

- Sometimes need to fail out of PoP's
  - 3rd party provider issues (e.g. transit links going down)
  - Infrastructure maintenance
- Withdraw anycast route announcements
- Fail healthchecks on proxy to drain unicast traffic